Virology **279,** 109–115 (2001)
doi:10.1006/viro.2000.0716, available online at http://www.idealibrary.com on IDEAL®

Complete Nucleotide Sequence and Analysis of a Novel Human Papillomavirus (HPV 84) Genome Cloned by an Overlapping PCR Method

Masanori Terai*,† and Robert D. Burk*,†,‡,[1]

*Department of Microbiology and Immunology, †Department of Pediatrics, and ‡Department of Epidemiology and Social Medicine, Comprehensive Cancer Center, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461

Received August 22, 2000; returned to author for revision October 3, 2000; accepted October 13, 2000

Molecular diagnosis of human papillomaviruses (HPVs) in cervicovaginal samples reveals a plethora of known and novel HPV genomes. We describe the use of an overlapping PCR method to clone and analyze the complete genome of HPV 84 from cervicovaginal cells obtained from a 21-year-old Caucasian female with a normal Pap smear. The 7948-bp complete nucleotide sequence of HPV 84 was determined from five overlapping PCR products by sequence walking. A BLAST homology search demonstrated that HPV 84 was most closely related to HPV 61 (89%), HPV 72 (86%), and HPV 83 (85%) by nucleotide sequence analysis of the L1 open reading frame, placing it in the HPV genome homology group A3. Previously, this virus had been identified as Pap155. Based on extensive epidemiological data, HPV 84 is a highly prevalent genital papillomavirus primarily detected in normal and HIV-infected women. © 2001 Academic Press

Key Words: HPV; human papillomavirus; cervical cancer; STD; sexually transmitted disease; overlapping PCR; phylogeny.

The papillomaviruses are a heterogeneous group of DNA viruses with circular double-stranded DNA genomes about 8 kb in size. These viruses infect humans as well as numerous and diverse animal species. Human papillomaviruses (HPVs) are classified into more than 100 types to date, including 84 that have been cloned and characterized, whereas others have been identified by partial sequence analysis of a polymerase chain reaction (PCR) product. For a papillomavirus to be recognized as a distinct type, the full genome should be cloned and the sequence of the L1 gene should be no more than 90% similar to previously typed HPVs (Delius *et al.,* 1998). Most genital benign and malignant neoplastic lesions are associated with HPV infections. HPVs are categorized into mucosal and cutaneous types, as well as epidermodysplasia verruciformis (EV)-associated types. In addition, HPV types are also categorized into low- and high-risk types in accordance with their association with malignant lesions (Bosch *et al.,* 1995; Van Ranst *et al.,* 1996).

We have characterized a novel genital HPV type, now called HPV 84, that was amplified and cloned from the cervicovaginal cells obtained from a 21-year-old Caucasian female with a normal cervical cytologic analysis by use of an overlapping PCR method. An overlapping PCR method has been reported as a useful tool for obtaining HPV DNA from specimens with low copy numbers (Forslund and Hansson, 1996). The genome of HPV 84 contains the 452-bp MY09/MY11 DNA fragment previously known as either MM8 or Pap155 (Manos *et al.,* 1994). This partial genome was originally characterized from a vulvar sample from a patient with genital warts and cervical abnormalities (Manos *et al.,* 1994).

In this communication, we report the cloning of Pap155 using an overlapping PCR amplification method and the determination of the complete nucleotide sequence. In addition, phylogenetic analysis has been conducted and the structure of the predicted HPV 84 proteins has been determined.

RESULTS AND DISCUSSION

Complete nucleotide sequence and predicted open reading frames

The PCR primers specific for the HPV 84 sequence are shown in Table 1 and the organization of the five overlapping PCR products is shown in Fig. 1. The assembled sequence of the viral genome revealed a total size of 7948 bp with a G+C content of 46.44%. The complete nucleotide sequence is shown in Fig. 2. The sequence is available in GenBank under Accession No. AF293960. The DNA clones and sequence were submitted to the Human Papillomavirus Reference Laboratory (Heidelberg), and the virus was assigned the number HPV 84. A basic local alignment sequence tool (BLAST) homology search using nucleotide sequence of the L1 open reading frame (ORF) showed that HPV 84 is most closely related to HPV 61 (89% homology), HPV 72 (86% homology), and HPV 83 (85% homology), thus, satisfying the criteria for a new HPV type (Delius *et al.,* 1998). This

---

[1] To whom correspondence and reprint requests should be addressed. Fax: (718) 430-8975. E-mail: burk@aecom.yu.edu.

0042-6822/01 $35.00

TABLE 1

Primers Used to Clone Overlapping DNA Fragments of HPV 84

Plasmid No.[a]	Insert size (kb)	Primer sequence (primer name; primer direction)
1	0.7	5′-TGA CCC CGA GGA AAC AAT TA-3′ (5300F; forward)
		5′-GCA GAT ATA CCT TGC CGT CA-3′ (SP84R; reverse)
2	4.5	5′-AAA CAG GAT TAC CAC TTG CA-3′ (500Fout; forward)
		5′-CCT ACG TGC AGT TAA CGC CG-3′ (5300Rout; reverse)
3	1.0	5′-GGG TAG GTC TAC TGT TTC CT-3′ (MY98F; forward)
		5′-TGC AGC ACT TCC CTG TCC ACT GTC C-3′ (500R; reverse)
4	1.8	5′-ACG TGT GCC CTA TTC TTT TTC A-3′ (MY84; forward)
		5′-AGG AAA CAG TAG ACC TAC CCA G-3′ (MY89; reverse)
5	0.6	5′-TAT GTG TCT GTT TGT TTG GTG-3′ (Pap155-A; forward)
		5′-ACT ATG CAC TTA AGG GAA AG-3′ (Pap155-B; reverse)

[a] See Fig. 1.

relationship places HPV 84 in the papillomavirus homology group A3 based on the Human Papillomaviruses 1997 Compendium (http://www.stdgen.lanl.gov/stdgen/virus/hpv/compendium/htdocs/COMPENDIUM_PDF/97PDF/1/Intro97.pdf). HPV 84 represents the previously partially characterized HPV known as either MM8 or Pap155 based on the 452-bp DNA fragment amplified by the MY09/MY11 consensus PCR primers (Manos *et al.,* 1994). This partial genome was not detected with the GP5/GP6 set of primers or with any of the JC9813-specific primers (Feoli-Fonseca *et al.,* 1998).

Examination of the HPV 84 sequence for potential genes showed the typical complement of papillomavirus ORFs. The predicted ORFs are summarized in Table 2A. Table 2B shows the homology of putative HPV 84 proteins to the analogous proteins of several closely related HPV types. HPV 84 proteins were closely related to HPV 61, HPV 72, and HPV 83 proteins. Interestingly, the E6, E2, and E4 ORFs were most closely related to those of HPV 61, whereas the E7, E1, L2, and L1 ORFs were closest to HPV 83. The presence of an E5 protein situated between the end of the E2 ORF and the start of the L2 ORF, which is found in some but not all HPVs, was sought by comparison of this region in HPV 84 to the complete papillomavirus database. None of the small ORFs in this region of HPV 84 showed significant homology with known E5 proteins.

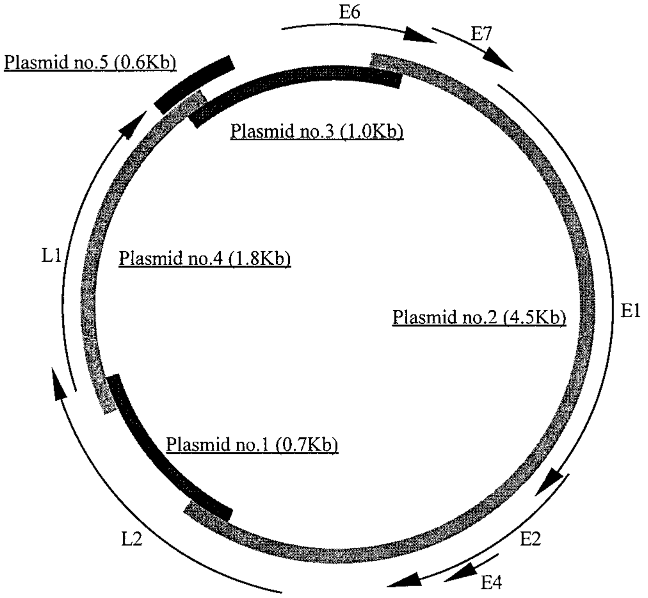

FIG. 1. Organization of the five cloned overlapping PCR products and location of predicted ORFs. (See Table 1 for sequence of primers and Table 2A for exact location of ORFs.)

Phylogenetic analysis

To investigate the relationship between HPV 84 and related HPV genomes, nucleotide and amino acid sequences of the ORFs in HPV 84 were aligned with the corresponding sequences of HPVs from group A3. Sequences were aligned using Sequencer software and verified manually. Since not all related HPV genomes were completely sequenced, a set of trees was constructed based on the sequence of the HPV L1 region amplified by MY09/MY11. A representative tree is shown in Fig. 3. The tree demonstrated that HPV 84 was most closely related to L1AE6 (Ho *et al.,* 1998c), also known as CP6108 (Peyton and Wheeler, 1994), and other members of the A3 clade.

The LCR

The sequence between the end of the L1 ORF and the beginning of the E6 ORF is called the long control region (LCR), also known as the upstream regulatory region (URR) or noncoding region (NCR). This region contains many of the *cis*-acting regulatory sequences that control viral transcription and replication. The LCRs of related HPV types, HPV 61, HPV 72, HPV 83, and HPV 16, are 776, 755, 851, and 814 bp long, respectively. The LCR of HPV 84 is 772 bp long. One of the elements present in papillomavirus LCRs is the origin of DNA replication (ORI). The ORI typically contains an E1 binding site between

```
   1 ATGCCCAACG GACGCTACCA CCCCACCAAT ATTTTTGTGC TGTGCCAGGA ATACGAGGTG GAGTTCGACG ACCTACGATT AATTTGCATC TTTTGCAAGG
 101 AAGAATTAAC GGAAGGCGAA GTGCTGGCCT TTGCAGTAAA GGAATTACTA ATTGTTTGGA GGTATAATTT CCCTCATGGG GTGTGCATGA AATGCTTATG
 201 CAGGGAAGCC AAAGTACGTG AGCTACGCCA CTGGGATTAC TCCAGCTTTG GACCAACAGT GGAAGAAGAA ACAGGATTAC CACTTGCACA AATAAATATA
 301 AGGTGCCACG CGTGCTGCAA GCCATTGTGC TATCAGGAAA AGGAGTATAT GGTGGAATTG CAGTTGCTAT TCCACAAAAT AGCTGGACAG TGGACAGGGA
 401 AGTGCTGCAA CTGTAGGGTA ACATGCGCGG CCAGACGCCA ACGTTAAAGG ATATTATTTT AACAGATATA CCAGATGTAG TTAGTTTATA CTGTGACGAG
 501 CAATTGTTAG ACAGCTCAGA GGAGGAGGAT AATGGGGATT GCTTGCGTGA CCAACCTGCA AAACCAGCAC AGCTGGCCTA TAGGGTGGTA ACAGAGTGTG
 601 GGTTGTGTAG CCGTCCAGTT AGGCTGGCGG TGCTTTGTGG AGTGGAAGAC CTCAGAAAGC TGCAGCAGCT TATAGTGGAA GCAGTGGCCA TAGTGTGTCC
 701 CGGCTGTGCA TAATATGGCA GAGTCACCTG AAGGTACCGA TGGGGGGACT GGCGGGGGGG CTGGGGGATG GTTTGTGGTG GAAGCCATGG TACACCAGGC
 801 AAACACCGAT GTGTCTAGCG ATGAGGACGA AACACGTATA GATACAGGAG AAGATTTGGT AGACTTTATA GATGATACCA GACATCCAGG GGATGGACAG
 901 GAAGTGCCGT TAGATTTGTA TGTTCAACAA ACAATACAGG ATGACGCTGC AACGGTGCAA GCCCTAAAAC GAAAGTTTAT GGGAAGCCCA GCAACCAGTT
1001 CCTGCGCATC CTGGGTGGAT AATGAACTAA GTCCACGGTT GAATGCAATA TGTTTGCAGA AAAGGCACGA CAAGGCCAGG AGAAGGTTGT TTGACCAGGA
1101 CAGTGGGTAT GGCAATACGC AGGTGGATAT CGGAACATCC GAGAGTCAGG TACCAGGGGG AACCACTAAC AGCGGGGGGG GAGACGTAAC ACAGGACGTT
1201 GTAGAGGAGG AACGCAGGGG GGGGGATGGG GAAACACAGC CACATGCAAA CACGCAGCAC ACACAGCAGA CGGAAGGCAC ACCAGACGCA TTAGACCTGC
1301 TGCAGGTTAG TAACCTAAGG GTAAAACTAT TAGGAAAATT TAAGGAACTA TTTGGGCTGT CCTTTATGGA TCTGGTAAGA CAATTTAAAA GTAATAAGTC
1401 AACATGTGGA GACTGGGTGG TGGGGGCATT CGGGGTGTAC CATGCAGTGG CAGAAGCAGC AAAAACACTG CTACAGCCTG TGTGTGAGTA TGCACATATC
1501 CAAACATTAA CTAGTGAATG GGGAATGGTA ATGCTGCTGC TACTGCGCTT TAAGTGTAAC AAAAGCAGGG AAACAGTAGC GCACTGTATA GGTGGCATTT
1601 TAAATGTACC CGAAAAACGT ATGCTAATAG AGCCCCCAAA GCAACGCAGC GGGCCATGTG CACTATATTG GTATAGAACA GCAATGGGCA ATGCATGTGA
1701 AGTGTTAGGG GAAACACCAG ACTGGATAGT ACGACAAACT GTAATTGGGC ATGCAATGGG GGAAACGCAG TTTAGTTTAT CAAAACTAGT GCAGTGGGCA
1801 TATGACAATG ACATTACAGA TGAAAGTGAA CTGGCCTATG AATATGCACA ATTGGGAACA GAGGAGCCAA ATGCAGCCGC CTTCTTAGCA AGCAATTGTC
1901 AGGCACGATA TATAAAGGAT GCAATGATAA TGTGCAGACA TTATAGACGT GCAGAACAGA CACGTATGAG TATGTCACAA TGGATAACAT ATAGGGGGCG
2001 CAAGGTAGCT GATACAGGCG ATTGGAGGCA TATAGTAAAA CTATTAAGAT ATCAAGGAAT AGAATTTATT AGCTTTATGA CAGCATTAAA GCAATTCCTA
2101 AAGGGCACAC CAAAAAAAGG CTGTTTGGTA TTTTATGGAC CCAGCGATAC AGGCAAATCG CTGTTTTGCA TGAGTTTAAT AAACTATTTA GGTGGAACAG
2201 TTATATCCTT TGTAAATTCT ACTAGTCATT TTTGGCTGTC GCCACTGGCA GATGCTAAAA TAGGATTGCT AGACGACGCC ACATATCAAT GCTGGATATA
2301 TATGGATACA TATTTAAGAA GTGTGCTAGA TGGCAATGTA ATAAGCGTAG ACAGAAAACA TAAAAACCTA GTACAGTTAA AATGTCCTCC GTTACTAATA
2401 ACAACAAACA TAAATCCAGA AACTGATGAC ACATTTAAAT ACTTACGCAG CCGAATGGTT ATTTTTCCCT TTTTAAACAA GTGCCCACTG GATGCCAATG
2501 GAGACCCAGT GTATCAATTA AATAATGAAA ATTGGAAATC CTTTTTTCGA AGGTCGTGGG CACGCTTAGA CTTAACCCAG GAGGAGGAGG AGGAGGAGGA
2601 GTCCCAAAAT GGAAACCCTA GCCGACCGTT TAGATGCGTG CCAGGAGAAG CTAATAGACC TTTATGAAAA AGATAGCAAC AAGCTTGAGG ACCAGTTAAT
2701 GCATTGGTAC TACACGCGCC TGGAACAAGC TATGCTGTTT AAGGCAAGAG AAGCAGGACT AAGACACATA GGCCACCAGG TGGTGCCCAC ACTTAGTGTG
2801 ACAAAAGAAA AAGCTCGGCA GGCCATAATG GTGCATTTAT CTTTGCAAAG CCTAAATAAT AGTGCATTTA AGCACGAGCC ATGGACATTG CAGGACACAT
2901 CATTGAACAT GTGGACAGTA GCCCCAAAAG GGTGTTGGAA GAAACATGGA CAGCCCATCA GAGTAAAATA TGATGGAGAA GATGATAAAG AAATGGAATA
3001 TGTAAACTGG GGGTTTATAT ATGTACATTG TGCCAGTGAG GACACCTGGT ATAAGGTACC TGGACAGATT AGCAACAGAG GGCTGTATTA TGAACTGCAA
3101 GGCTGCAAAC ATTACTATGT GGAATTTGCA AAGGAGGCAA AACAGTATGG GGTAAAAAAC ATATGGGAGG TGTATATGGG AGGCAAAATA ATTTACCATG
3201 CATGCGACTC TGTATCCAGC ACCCAGGACG CCGTGCCAGA AGTACCCACT GCTGAAACTG ATGCACAGGT GCACCACACC ACCACCGCGC CCACCCCCAC
3301 CACCCAGCGC TGCACCAAGA CTGCCCCCCA GGTGCAGGCG CCGCCTGCTA AGCGACAGCG ACTCGGAGGG GACACAGTTC AGCAGCCCGA CTCTACACAA
3401 GGACACCGGC CGGTTGACAG TTGCAGCACA AGGACAAACA ATAACTGTGA CAGCCCAAAA CGGCCACGGG AACACAGTAA CTGTGACAGT GCACCTGTCC
3501 TACACCTAAA AGGTCAATCC AATAGCCTTA AGTGCTTTCG ATATAGGTTA CACCAGTCGG TGCCTGACCT GTTTGAAAGG GCATCGTCCA CGTGGAAGTG
3601 GACCTGCGGG GGGGAGGGCG ACAAAACATC ATATGTAACA CTATGGTATA AAAGTACGGA CCAACGCAAA CAGCTCTTGG CACGTGTACA TATACCAAAA
3701 GGCATTGTAG CCACACTTGG TAGTATGTCT ATGGTTATGT AAACAGTACC CATTGTATAT GTATAGTGTA AATGCTGTAT TACTGAGCCA TTTGTAACAA
3801 TTGTGTATGT CCTACATACC TGGCAAATAG GGCACAACCC TGGAACCTAT TGTGTGGGCA GTGCATGTAT GGTGTAATGT GTTGCTGCTG CTAGTTCTTT
3901 TCTGGCTTTC CCACCTATCT GCTTTTGTTG CATTTGTTGT GTTTGTGTGT GTTTTATATT TTGGGTTGCT GCTGTTATAT TTGCAGGTGC TGTGGTACAT
4001 TGAATTGTTA TAACCACACA ACCAGCCAAC GACTGCTGCT ACATATCCCT GTACAGTCCC GGTCCTACCT ATTATGTATC CCATGCAATT ACGGGGTCCC
4101 CAGGGTGGGT ATGACATAGT GGTGTTTGAG CGTGGTGATG TAGGTCTTTT TACTATTATT TTGCTTTTAA TTGTTATTAT TTTACTTTGG CTGTGCCACC
4201 GTGTTATGCA GTTTTAAAGT GCTTTTGTAT TTTTGTACAT ACTATAATAA ACACTTGGTA CCATGCCCAA GGTTCTTAAA CGTCGCAAGC GTGCTTCTGC
4301 CACGGACCTT TATCGCACCT GCAAGGCCAC TGGAACACGT CCTGCTGATG TTATTCCTAA GGTGGAAGGG GATACATGGG CGGATCGCTT TTTAAAATGG
4401 GCCAGCCTGG GTGTGTTTTT TGGTGGTCTG GGTATTGGCA CATCCTCTGG CACGGGTGGC CGTACGGGCT ATATACCCCT TGGTACCCGG CCTCCTACTG
4501 TGGTTGATGT AGGCCCTACG GCCCGTCCGC CCGTTGTCAT TGAACCCGTT GGCGCTGCAG ACCCTTCAAT TGTTACCTTG GTGGAGGATT CCAGTGTTAT
4601 TAATGCCGGG GCCCCTTTTC CTAACTTTAC GGGTACTGGT GGGTTTGAGG TCACCACATC CTCTATTACC ACACCTGCGG TTTTAGACAT CACCCCCTCG
4701 GGGTCGTCTG TGCAGGTCAG TAGTACCAGT TATGTTAACC CGCTGTTTAC TGAGCCTTCT ATTATTGAGC CTCCCCAGGC AGGGGACATT ACTGGTCATG
4801 TGTTGTCTAG TACAGCCACG TCCGGGTCAC ATACCTACGA AGAAATCCCC ATGCAAACCT TTGCCGTGCA GGGGGGTACG GGCCTAGAAC CCATAAGTAG
4901 TACACCCACA CCTGGGGTGC GGCGCCTTGC AGGGCCTCGC TTAAATTTAT ACAGTCGTGC CACTCAACAG GTTCCTGTGG CTGACACTGC ATTCTTATCA
5001 CGTCCTGAAT CATTTGTTAC CTTTGACAAC CCTGTGTTTG ACCCCGAGGA AACAATTATA TTTGAACATC CTAGTTTACA TGCGCCACCC GATCCTGACT
5101 TTCTGGATAT TGTCACTTTA CATAGGCCGG CGTTAACTGC ACGTAGGTCA GGTGTACGGT TCAGTCGCAT AGGCCAAAGG GCGTCTATGC GCACACGCAG
5201 TGGTAAACAT ATTGGGGCCC GTGTTCATTT TTATCATGAC CTTAGCCCTA TACCACATCT TGAGGATATT GAGCTGCAGC CCCTGGTATC ATCCTCTGCT
5301 GTGCCTACGG ATTCATTATA TGATATATAT GCAGATGATG CCCATTTGTC CTCTGTATTG CGTCCCCCTT CCGTTTCTGC CCTACGTCCA GCTTCCCCCT
5401 TTGCCTCTGC AGACCTTTCT GCCACCTCCA TAACCGCCTC CACATATGAT AATGTCACTG TCCCGTTGTT TTCTGGCACC GATGTGCCCG TCTATACAGG
5501 CCCTGATATT GACCATTCTG CTGCTCCCTC CGCACCTCCC TTTGTTCCTG TTATTCCTAG CACAACCCCA TATGCTATTT ATATCCTGGG GTCAGATTAT
5601 TATTTGCTTC CTAATTATAT ATTTTTTCCT AAAAAGCGTA AACGTGTGCC CTATTCTTTT TCAGATGGCT TTGTGGCGGC CTGGTGACGG CAAGGTATAT
5701 CTGCCCCCCA CTCCCGTGTC TAAGGTTATC AGCACGGATC GTTATGTCTC TCGCACCAAC TTATTTTATT ATGGTGGTAG TTCTCGCCTG CTTACTGTGG
5801 GACATCCATA TTATTCTGTT CCTGTGTCTA CCCCTGGGCA AAACAACAAA AAGGCCACTA TCCCCAAGGT TTCTGGGTAT CAATACAGGG TGTTTAGGGT
5901 CCATTTACCT GACCCCAATA AGTTTGGTCT TCCGGATGCA CAATTATATA ATCCTGACAC CGAGCGCCTT GTGTGGGCTT GTAGGGGTGT TGAGGTAGGC
6001 CGCGGGCAGC CTTTAGGCGT TGGCACTAGT GGCCACCCTT TATACAATAG GCTTGATGAC ACTGAAAACA CCCCTTTTCT TGCTGCTGGG GACACTGACA
6101 GTAGGGATAA TGTTTCTGTG GATTATAAAC AAACACAGCT GTTAATTATA GGCTGTAAGC CGTCTATTGG GGAACATTGG GGTAAGGGTA CTGTATGTAC
6201 TAATGTGCAA TATCGTGCGG GTGATTGCCC CCCTTTGCAG TTCACTAATT CTACCATTGA GGATGGTGAT ATGGTTGAGG CTGGCTATGG TGCTATTGAT
6301 TTTGCCACAC TGCAGGAAAG TAAGTCAGAG GTGCCTTTGG ATCTCTGCAC CACTACCTGT AAATATCCTG ATTATTTACA AATGGCTGCA GAGCCGTATG
6401 GCGATTGTAT GTTTTTTTGT CTTCGGCGGG AGCAAATGTT TGCTAGGCAT TTTTTTAACA GGCAGGGTAC TATGGGTGAG GAGGTGCCTC AGTCTTTTTA
6501 CCTTAAGGGG ACCTCCTCCC GGGCAACCCT TAGCAGTTCT GTGTATGCCC CTACACCCAG TGGCTCTATG GTGTCCTCGG ATTCCCAATT GTTTAATAAG
6601 CCATACTGGT TGCAAAAGGC CCAGGGTCAT AACAATGGTA TATGCTGGTT TAATCAATTG TTTGTCACGG TGGTAGATAC CACCCGCAGC ACCAATTTTA
6701 CTATTAGTGC TGCTACCAAC ACCGAATCAG AATATAAACC TACCAATTTT AAGGAATACC TAAGACATGT GGAGGAATAT GATTTGCAGT TTATATTCCA
6801 GTTGTGTAAG GTCCGTCTGA CTCCAGAGGT CATGTCCTAT TTACATACTA TGAATGACTC CTTATTAGAT GAGTGGAATT TTGGTGTTGT GCCCCCTCCC
6901 TCCACAAGTT TAGATGATAC CTATAGGTAC TTGCAGTCTC GCGCCATTAC TTGCCAAAAG GGGGCCGCCG CCGCCAAGCC TAAGGACGAT CCTTATGCTG
7001 GCATGTCCTT TTGGGATGTA GATTTAAAGG ACAAGTTTTC TACTGATTTG GATCAGTATC CCTTGGGTCG CAAGTTTTTA TTGCAGTCTG CCCCACGTTC
7101 CACCCTGGTG TCCCGTAAAC GTACAGCGTC TGCCTCTACC CCCCCTGCCT CCAAACGGCG TAAGGCCAAA AAGTAATATG TGTCTGTTTG TTTGGTGTCT
7201 GTGTCTGTTT GTTTGGTGTC TGTACAATCT CTGTGTGTGC CATTTTGTAT GTGTTTTGTA CTGTTCTGTA TGTGCCCTGT TATGTATAAT GTGTATGTAT
7301 TTATGGAATG CGTGCCCCTG TATGTTGTGT ATGTTGTGTA CATGTATTTG TGGAATGTGT GTCATGTTAT GGATCAATAA ATTGTGTGTC ATTGTGGGTT
7401 TCATGTCCGG CTGCACCCTG TGAGTAAGTG TGCACCATGT ACACGCCTGG GTAGGTCTAC TGTTTCCTTT AAGGGTATAA AGCCTCCATT TTGTATGCAA
7501 CCGTTTTCGG TCTCCCGCCT TTTCGGTCTG TGGTATGGCA CTGTCCCAGG TACAGCTAAT CCTTTGGCAG CCCAACATCC TGCGTAGGCA GCTGCAGACG
7601 TCCGCACCCA GGTGTGTCTG TCAGCTATTT CAATATGCTA ATAAATGTAT TGTCTACAGT TTATAGGTTT ACTTACTCAT CTGCAAAAAA TATGCTTTTA
7701 GGCAGTTTTT TGGCCTACAA CTTTCCCTTA AGTGCATAGT TGGCAGGCCG TGCACACTGT GCTGCCAGGT TTCTGTCTTG TAAAATACCA CAATACTTAT
7801 GTAACACACC CGTTTCGGTT GCTATGTTAT TCATACTATT TTCTTATACT TTACAACAAT GATACAGTAA AAAAATAGGG AGGGACCGAA TTCGGTCCCA
7901 CCGAAAGGGA TACATATATA AAGGGCAGCA AACGGTACCC GGACAGCC
```

FIG. 2. Complete nucleotide sequence of HPV 84.

TABLE 2A

Location of Predicted ORFs and Size of Putative Proteins

ORF	Start position	First ATG	Stop codon	Length of protein-coding sequence (bp)	Amino acids	Predicted molecular mass of protein (kDa)
E6	7922	1	447	444	148	17.4
E7	417	423	713	288	96	10.5
E1	694	715	2667	1950	650	72.8
E2	2576	2609	3742	1131	377	43.0
E4	3192	3198	3509	309	103	11.2
L2	4251	4263	5687	1422	474	50.6
L1	5434	5665	7176	1509	503	56.4

the putative binding sites for two E2 binding sites (ACCN$_6$GGT) (Lu *et al.,* 1993; Sun *et al.,* 1996).

The HPV 84 LCR contains three E2 binding sites (see Fig. 4). Two sites at 7500 and 7896 contain E1 binding sites at positions 7841 and 7860. This region probably represents the HPV 84 origin of DNA replication. Papillomavirus LCRs also contain multiple binding sites for transcriptional regulatory factors such as AP-1 (Chan *et al.,* 1990), NF-1 (Apt *et al.,* 1993), SP-1 (Gloss and Bernard, 1990), transcriptional enhancer factor (TEF)-1 (Ishiji *et al.,* 1992), and YY-1 (Dong *et al.,* 1994; May *et al.,* 1994), among others. The organization of the HPV 84 LCR is consistent with that of other genital papillomaviruses. Regulatory sites include a TATA box representing the E6/E7 promoter at positions 7915 and 7921, just one base upstream from the start of the E6 ORF at position 7922. The predicted locations of these sites within the LCR of HPV 84-related HPV types, HPV 61, HPV 72, and HPV 83, and HPV 16 LCR are shown in Fig. 4. Based on the LCR length and position of multiple binding sites, the region of HPV 84 is more similar to the low-risk HPV types.

Diversity of HPV 84

To better understand the evolution and generation of independent HPV types, we have compared the amino acid and nucleotide differences of HPV 84 and those of HPV 61, the most closely related HPV type, as shown in Table 3. Both the nucleotide sequence and the amino acid sequence of the L1 ORF are the most highly conserved. In contrast, the E7 and E2 ORFs have highly conserved nucleotide sequences, but more divergent amino acid sequences. Most of the nucleotide substitutions within each ORF were found in the third position (42.2 to 60.3%). Approximately 37 to 52% of all codons analyzed had a nucleotide change in the third position. The conservation of DNA sequence and the lack of a continuum between HPV 84 and HPV 61, the most closely related type, indicate that HPV 84 and HPV 61 have emerged as independent entities, equivalent to "species" (Bernard, 1994; Van Ranst *et al.,* 1992). The distribution of nucleotide changes in first, second, and third positions suggests a gradual mode of molecular evolution of HPV 84 and HPV 61.

HPV 84 is of significant interest because it is one of the most common types found in the genital tract of women with normal cytology. For instance, HPV 84 was found in 7.0% of HPV-infected college women (Burk *et al.,* 1996; Ho *et al.,* 1998c). HPV 84 has also been detected in multiple HPV-type infections associated with condylomata acuminata occurring in immunosuppressed patients (Brown *et al.,* 1999). However, HPV 84 is uncommon in patients with dysplasia and cancer (Ho *et al.,* 1998a,b; Meyer *et al.,* 1998). One large study demonstrated that the proportion of women with HPV 84 was

TABLE 2B

Homology of HPV 84 Amino Acid Sequences with Related HPVs

HPV 84	E6	E7	E1	E2	E4	L2	L1
HPV 61	65.3 (%)	56.3	71.6	59.9	63.7	69.8	79.6
HPV 72	63.3	52.5	72.2	59.7	62.1	70.3	79.6
HPV 83	60.1	59.8	73.0	57.1	53.8	73.0	80.0

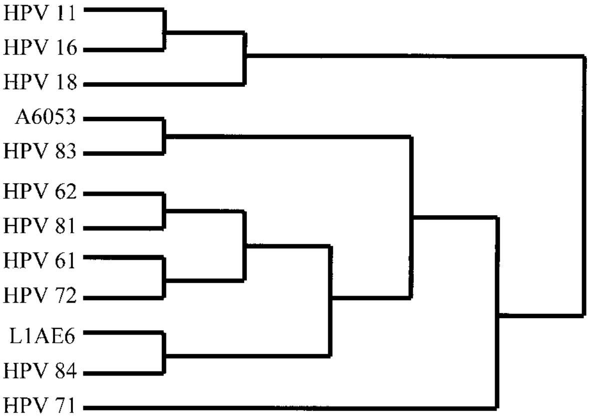

FIG. 3. Phylogenetic tree based on sequences of the HPV L1 region between the MY09/MY11 primers. Sequences were aligned and a tree was generated using the CLUSTAL program.

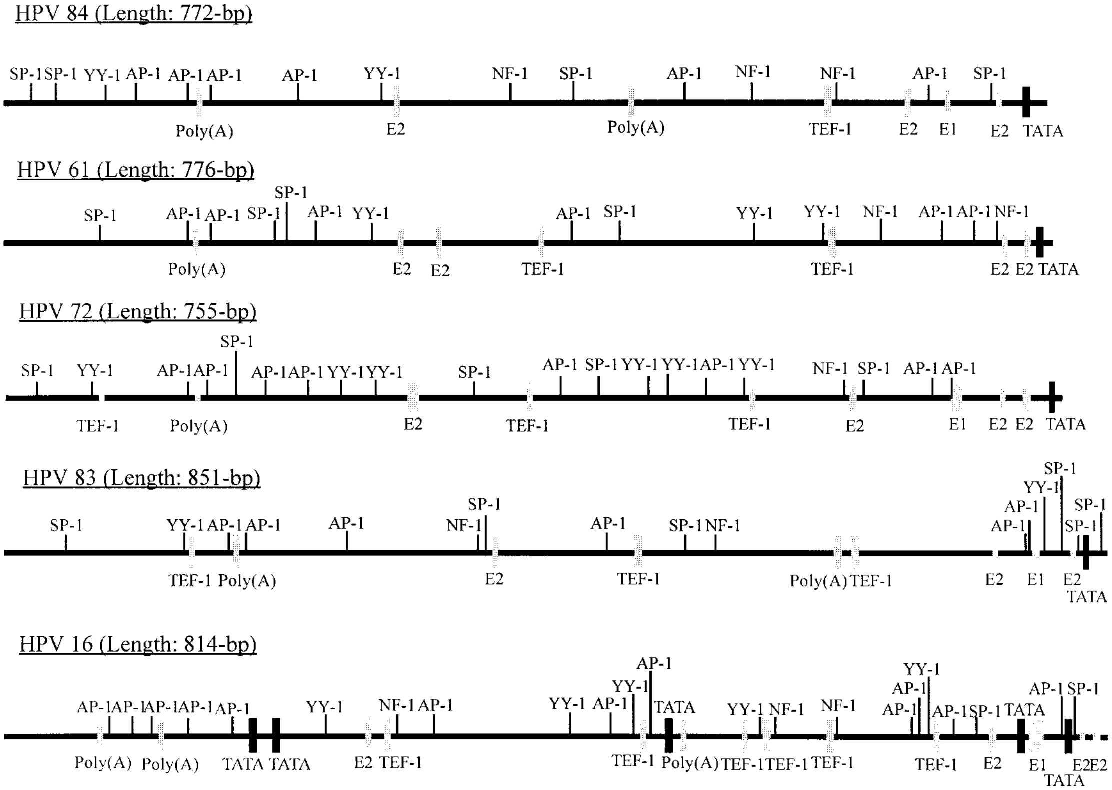

FIG. 4. Comparison HPV 84 LCR with related HPV types (HPV 61, HPV 72, and HPV 83) and HPV 16 LCR. LCR length and position of multiple binding sites (AP-1, NF-1, SP-1, TEF-1, YY-1, poly(A) signal, TATA box, and E1 and E2 binding domains) are shown, as indicated. TATA, TATA box (black box); TEF-1, transcriptional enhancer factor-1 (gray box); E1, E1 binding domain (gray box); E2, E2 binding domain (gray box); Poly(A), poly(A) signal (gray box).

higher in women with human immunodeficiency virus (HIV)-1 infection (4.4%) than in HIV-negative women (0.2%) and its prevalence increased inversely with the CD4 count in HIV-infected patients (Palefsky *et al.,* 1999). In addition, HPV 84 tends to occur as a multiple-type infection involving HPV 16 (Franco *et al.,* 1999). Thus, HPV 84 is a relatively prevalent genital tract papillomavirus in a variety of patient populations, including patients with HIV infection.

MATERIALS AND METHODS

Overlapping PCR

The HPV 84 genome was amplified by PCR as overlapping fragments that were cloned into a plasmid vector. Initial PCR primers were designed by alignment of closely related HPV genomes determined by phylogenetic analysis using the sequences of the 452-bp L1 region amplified by the MY09/MY11 consensus PCR

TABLE 3

Divergence between HPV 84 E6, E7, E1, E2, E4, E5, L2, and L1 Genes from Most Closely Related HPV Type (HPV 61)

Comparison	Codons compared[a]	Similarity (%) with compared codons[a]		Nucleotide substitutions by codon position[b] (%) (total: 100%)			Codons with a third position change[b] (%)
		Amino acid	Nucleotide	1st	2nd	3rd	
E6	144	65.3	68.1	32.6	15.2	52.2	50.0
E7	95	56.8	69.1	28.4	22.7	48.9	45.3
E1	645	72.4	73.5	26.7	17.9	55.4	44.0
E2	374	60.7	70.7	35.6	22.2	42.2	37.2
E4	97	67.0	70.8	28.2	21.2	50.6	44.3
L2	457	71.8	69.6	28.1	14.4	57.6	52.5
L1	501	80.0	75.8	24.8	14.9	60.3	43.7

[a] Not counting gaps and terminal extensions.

[b] Including silent codons.

primers (Chan *et al.,* 1995; Manos *et al.,* 1994). A clinical sample previously identified to contain a high copy number of Pap155 was used (Burk *et al.,* 1996; Ho *et al.,* 1998c). The Pap155-positive sample DNA was subjected to PCR amplifications using either Gold *Taq* DNA polymerase (Perkin–Elmer Applied Biosystems, Foster City, CA) or an equal mixture of Gold *Taq* (Perkin–Elmer) and *Pwo* DNA polymerase (Platinum *Taq* DNA Polymerase High Fidelity, Gibco BRL, Rockville, MD). *Pwo* polymerase has an inherent 3′–5′ exonuclease proofreading activity. The PCR products were separated by electrophoresis in agarose gels, stained with ethidium bromide, and visualized under an ultraviolet transilluminator. After confirmation of appropriate product size, each PCR product was purified (Qiagen Gel Extraction Kit, Qiagen, Valencia, CA) and ligated into the *p*-GEM-T Easy vector (Promega, Madison, WI) according to the manufacturer's instructions.

Sequencing strategy

To determine the nucleotide sequence, each DNA insert was sequenced using SP6 and T7 primers flanking the HPV insert. Additional primers were designed by walking through the insert using the most recent sequence (Delius and Hofmann, 1994). Sequencing was done on an ABI Prism Model 377 automated sequencer (Perkin–Elmer Applied Biosystems) in the Einstein core facility. The overlapping sequence fragments were assembled manually and confirmed by sequencing the complementary strand. Several additional primers were used to clarify sequence ambiguities. Once assembled, the sequence was analyzed for homology to other HPVs using BLAST software (Altschul *et al.,* 1997). The same software was used to determine protein sequence homologies.

Phylogenetic analysis

Phylogenetic trees for all published sequences are available from the Human Papillomaviruses 1997 Compendium on Line (http://www.stdgen.lanl.gov/stdgen/virus/hpv/compendium/htdocs/COMPENDIUM_PDF/97PDF/1/Intro97.pdf). Phylogenetic trees were calculated from individual ORFs, putative proteins, and LCRs, to determine the association of HPV 84 with the available HPV sequences available through the HPV database using public domain software (Higgins and Sharp, 1988).

ACKNOWLEDGMENTS

This work was supported in part by grants from the N.I.H. to R.D.B. Assignment of the HPV type number was kindly performed by Dr. Ethel-Michele de Villiers, Human Papillomavirus Reference Laboratory, Heidelberg, Germany.

REFERENCES

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**(17), 3389–3402.

Apt, D., Chong, T., Liu, Y., and Bernard, H. U. (1993). Nuclear factor I and epithelial cell-specific transcription of human papillomavirus type 16. *J. Virol.* **67**(8), 4455–4463.

Bernard, H. U. (1994). Coevolution of papillomaviruses with human populations. *Trends Microbiol.* **2**(4), 140–143.

Bosch, F. X., Manos, M. M., Munoz, N., Sherman, M., Jansen, A., Peto, J., Schiffman, M., Moreno, V., Kurman, R., Shah, K., and Group, I. S. (1995). Prevalence of human papillomavirus in cervical cancer: A worldwide perspective. *J. Natl. Cancer Inst.* **87,** 796–802.

Brown, D. R., Schroeder, J. M., Bryan, J. T., Stoler, M. H., and Fife, K. H. (1999). Detection of multiple human papillomavirus types in condylomata acuminata lesions from otherwise healthy and immunosuppressed patients. *J. Clin. Microbiol.* **37**(10), 3316–3322.

Burk, R. D., Ho, G. Y. F., Beardsley, L., Lempa, M., Peters, M., and Bierman, R. (1996). Sexual practices and partner selection are the predominant risk factors for genital HPV infection in young women. *J. Infect. Dis.* **174,** 679–689.

Chan, S.-Y., Delius, H., Halpern, A. L., and Bernard, H.-U. (1995). Analysis of genomic sequences of 95 papillomavirus types: Uniting typing, phylogeny, and taxonomy. *J. Virol.* **69,** 3074–3083.

Chan, W. K., Chong, T., Bernard, H. U., and Klock, G. (1990). Transcription of the transforming genes of the oncogenic human papillomavirus-16 is stimulated by tumor promoters through AP1 binding sites. *Nucleic Acids Res.* **18**(4), 763–769.

Delius, H., and Hofmann, B. (1994). Primer-directed sequencing of human papillomavirus types. *In* "Human Pathogenic Papillomaviruses" (H. zur Hausen, Ed.), Vol. 186, pp. 13–31. Springer-Verlag, Berlin.

Delius, H., Saegling, B., Bergmann, K., Shamanin, V., and De Villiers, E. M. (1998). The genomes of three of four novel HPV types, defined by differences of their L1 genes, show high conservation of the E7 gene and the URR. *Virology* **240,** 359–365.

Dong, X. P., Stubenrauch, F., Beyer-Finkler, E., and Pfister, H. (1994). Prevalence of deletions of YY1-binding sites in episomal HPV 16 DNA from cervical cancers. *Int. J. Cancer* **58**(6), 803–808.

Feoli-Fonseca, J. C., Oligny, L. L., Filion, M., Simard, P., Russo, P. A., and Yotov, W. V. (1998). JC9813—A putative novel human papillomavirus identified by PCR-DS. *Biochem. Biophys. Res. Commun.* **250,** 63–67.

Forslund, O., and Hansson, B. G. (1996). Human papillomavirus type 70 genome cloned from overlapping PCR products: Complete nucleotide sequence and genomic organization. *J. Clin. Microbiol.* **34**(4), 802–809.

Franco, E. L., Villa, L. L., Sobrinho, J. P., Prado, J. M., Rousseau, M. C., Desy, M., and Rohan, T. E. (1999). Epidemiology of acquisition and clearance of cervical human papillomavirus infection in women from a high-risk area for cervical cancer. *J. Infect. Dis.* **180**(5), 1415–1423.

Gloss, B., and Bernard, H. U. (1990). The E6/E7 promoter of human papillomavirus type 16 is activated in the absence of E2 proteins by a sequence-aberrant Sp1 distal element. *J. Virol.* **64**(11), 5577–5584.

Higgins, D. G., and Sharp, P. M. (1988). CLUSTAL: A package for performing multiple sequence alignment on a microcomputer. *Gene* **73**(1), 237–244.

Ho, G. Y., Kadish, A. S., Burk, R. D., Basu, J., Palan, P. R., Mikhail, M., and Romney, S. L. (1998a). HPV 16 and cigarette smoking as risk factors for high-grade cervical intra-epithelial neoplasia. *Int. J. Cancer* **78**(3), 281–285.

Ho, G. Y., Palan, P. R., Basu, J., Romney, S. L., Kadish, A. S., Mikhail, M., Wassertheil-Smoller, S., Runowicz, C., and Burk, R. D. (1998b). Viral characteristics of human papillomavirus infection and antioxidant levels as risk factors for cervical dysplasia. *Int. J. Cancer* **78**(5), 594–599.

Ho, G. Y. F., Bierman, R., Beardsley, L., Chang, C. J., and Burk, R. D. (1998c). Natural history of cervicovaginal papillomavirus infection in young women. *N. Engl. J. Med.* **338,** 423–428.

Ishiji, T., Lace, M. J., Parkkinen, S., Anderson, R. D., Haugen, T. H., Cripe,

T. P., Xiao, J. H., Davidson, I., Chambon, P., and Turek, L. P. (1992). Transcriptional enhancer factor (TEF)-1 and its cell-specific co-activator activate human papillomavirus-16 E6 and E7 oncogene transcription in keratinocytes and cervical carcinoma cells. *EMBO J.* **11**(6), 2271–2281.

Lu, J. Z., Sun, Y. N., Rose, R. C., Bonnez, W., and McCance, D. J. (1993). Two E2 binding sites (E2BS) alone or one E2BS plus an A/T-rich region are minimal requirements for the replication of the human papillomavirus type 11 origin. *J. Virol.* **67**(12), 7131–7139.

Manos, M. M., Waldman, J., Zhang, T. Y., Greer, C. E., Eichinger, G., Schiffman, M. H., and Wheeler, C. M. (1994). Epidemiology and partial nucleotide sequence of four novel genital human papillomaviruses. *J. Infect. Dis.* **170**(5), 1096–1099.

May, M., Dong, X. P., Beyer-Finkler, E., Stubenrauch, F., Fuchs, P. G., and Pfister, H. (1994). The E6/E7 promoter of extrachromosomal HPV16 DNA in cervical cancers escapes from cellular repression by mutation of target sequences for YY1. *EMBO J.* **13**(6), 1460–1466.

Meyer, T., Arndt, R., Christophers, E., Beckmann, E. R., Schroder, S., Gissmann, L., and Stockfleth, E. (1998). Association of rare human papillomavirus types with genital premalignant and malignant lesions. *J. Infect. Dis.* **178**(1), 252–255.

Palefsky, J. M., Minkoff, H., Kalish, L. A., Levine, A., Sacks, H. S., Garcia, P., Young, M., Melnick, S., Miotti, P., and Burk, R. (1999). Cervicovaginal human papillomavirus infection in human immunodeficiency virus-1 (HIV)-positive and high-risk HIV-negative women. *J. Natl. Cancer Inst.* **91**(3), 226–236.

Peyton, C. L., and Wheeler, C. M. (1994). Identification of five novel human papillomavirus sequences in the New Mexico triethnic population. *J. Infect. Dis.* **170**(5), 1089–1092.

Sun, Y. N., Lu, J. Z., and McCance, D. J. (1996). Mapping of HPV-11 E1 binding site and determination of other important cis elements for replication of the origin. *Virology* **216,** 219–222.

Van Ranst, M., Kaplan, J. B., and Burk, R. D. (1992). Phylogenetic classification of human papillomaviruses: Correlation with clinical manifestations. *J. Gen. Virol.* **73,** 2653–2660.

Van Ranst, M., Tachezy, R., and Burk, R. D. (1996). Human papillomaviruses: A never ending story? *In* “Papillomavirus Reviews: Current Research on Papillomaviruses” (C. Lacey, Ed.), pp. 1–20. Leeds Univ. Press, Leeds, UK.